Navigating Recorded Meetings with Content-Based Indices

Heather Richter
Department of Software and Information Systems
University of North Carolina at Charlotte
9201 University City Blvd. Charlotte, NC 28223
richter@uncc.edu

Gregory D. Abowd
College of Computing & GVU Center
Georgia Institute of Technology
Atlanta, GA 30332
abowd@cc.gatech.edu

ABSTRACT
Researchers have created a number of meeting capture applications in the past decade, yet relatively little research has focused on the review and use of captured meeting information. In this paper, we describe a controlled study of a mature meeting capture and access application, in which we observed subjects finding information within their own group meetings. The results demonstrate the importance of indices into the meeting record, especially those related to meeting content, and reveal a number of navigational behaviors with implications on the design of meeting capture and access applications and interfaces.

Author Keywords
Meeting capture and access, indexing, ubiquitous computing, evaluation.

ACM Classification Keywords
H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

INTRODUCTION
There have been a number of prototype meeting capture applications presented within the HCI community in the past decade, such as Distributed Meetings [4], Tivoli [9], and numerous others [3, 8, 14, 15], each exploring various ways of recording meeting discussions and outcomes. There are also a number of research efforts in the multimedia, audio and video perception communities that are investigating methods of recording and understanding meeting activities [1, 12]. Despite the plethora of meeting capture applications and technologies, relatively little work has focused on how recorded meeting information can be reviewed and utilized. Thus, we still do not understand how meeting capture will impact users.

Our goal is to advance that understanding through studies of the review, or access, of recorded meeting information. We have implemented and deployed a meeting capture prototype called TeamSpace [11]. Within TeamSpace, we emphasize capturing meeting-related artifacts and indices in order to facilitate meeting review, hypothesizing that indices based on such artifacts will provide valuable content-based browsing of the meeting record. In this paper, we describe a study on the interaction of users of TeamSpace in reviewing meeting information. Our results show how various indices into the captured stream were used, and emphasize the importance of the artifacts as content-related indices to facilitate finding a segment of interest. Additionally, we observed a number of low-level navigation behaviors that should be considered in any access application. These findings imply that capture systems should pay particular attention to capturing or recognizing content and content-related indices, and that there currently exist ways to do so without significant cost or technology.

In the follow section, we discuss related meeting capture prototypes and several evaluations of the use of captured material. Next, we introduce TeamSpace and detail the review interface. We then describe the study and the data collected, and discuss the users' interaction behaviors, introducing six navigational behaviors we observed. We further discuss the use of indices in TeamSpace and summarize the design implications of all our findings. Finally, we conclude with a general discussion of the evaluation and our future directions.

RELATED WORK
Researchers have explored numerous prototype applications for capturing meetings and presentations. All explore various ways to link, or index, the continuous audio or video streams with other information. Early work in meeting capture explored augmenting personal notes with audio, such as Filochat [14], and evolved to include meeting room video, such as in Notelook [3]. Others explored pen-based interactions by capturing whiteboard activity, such as Tivoli [9]; and others explored creating indices into the audio or video recording using audio, video,

and speech analysis techniques. For example, Jabber [8] used WordNet and lexical analysis to create indices of the discussion based on discussion topics. The Interactive Systems Lab at Carnegie Mellon are investigating a variety of multimodal techniques to perform speech recognition and participant identification, dialogue analysis, meeting summarization, recognize action items, and detect focus of attention and various speaker properties [1, 12].

Despite over a decade of research in capture and access, we have seen relatively little use of capture and access prototypes. The evaluation of eClass is the most extensive evaluation of a capture and access system to date [2]. Analysis of years of use provided a very comprehensive understanding of the navigation and adoption of the system. Closer to the meeting domain, He et al. [6, 7] have studied review of video summaries of informational presentations within Microsoft. In one study [6], they evaluated the extensive usage data from Microsoft's recorded presentations to examine real playback usage patterns. They found that in general, the number of viewers of any individual presentation segment decreases as time progresses. In other words, users tend to start watching a presentation at the beginning and stop sometime before the end. Additionally, the number of viewers spikes around slide changes, implying that users are navigating based on slides.

Most meeting capture and access systems built have been evaluated on a small scale, showing their usability and unobtrusiveness during the meeting. Initial studies suggested that a number of semantic cues can aid people in finding information in meeting recordings [15], and that using recorded meeting video can improve responses to questions about the meeting [8]. Wittaker et al. [14] observed users of Filochat and compared with traditional methods. Filochat gave participants the most confidence and the most accuracy in answering questions, but at the expense of added time. Thus, there seems to be a trade-off between access time and accuracy of recall.

Stifelman et al. performed a longitudinal field study with the Audio Notebook [13], observing four students and two reporters performing real tasks over a five-month period. The study showed that users had different ways of using the notebook as some augmented their existing notetaking, while others changed behavior and relied more heavily on the audio. This implies that capture and access services need to support multiple strategies of use for different people and tasks.

An extensive study of the use of meeting capture and access is from Moran et al. [9] who observed one person using Tivoli to aid in writing reports of intellectual property meetings. They identified several listening profiles, StraightThrough, Stop/Start, Re-listen, Skip-Ahead, and Non-Sequential, which changed over time. The user initially listened in a mostly linear fashion to the recording, but later was more focused, listening to only small portions of the audio, which he often transcribed and put directly into the report. Moran found that besides just reviewing the recording, the user needed a workspace to collect and organize relevant information from multiple sources. This study demonstrates the evolution of behavior that occurs as a user experiences and adapts to capture and access services. This only underscores the need to encourage repeated usage of capture services to enable many more long-term evaluations.

TEAMSPACE

TeamSpace is a prototype team collaborative workspace with integrated meeting capture and access capabilities [11]. Our goal was to relate the recorded material to the work that people do through the use of relevant artifacts. Thus, TeamSpace not only records the audio and video of a meeting, but also captures some of the artifacts that people use during those meetings and uses those interactions as indices into the recording. We began with simple meeting-related artifacts, namely agendas, action items, presentations, and participant attendance. Users can create, manage, and review meetings and their artifacts in the Web-based TeamSpace system, launching the MeetingClient application to capture distributed meetings and the MeetingViewer applet to review recordings.

We wanted to require as little instrumentation as possible to enable more locations for capture. The capture system was designed with an electronic whiteboard in mind, and desktop computers for distributed members, but can be run using just one laptop and an inexpensive microphone if desired. We capture events through users explicitly interacting with the artifacts contained in our software. The MeetingClient application automatically records audio and low-bandwidth video if desired. Additionally, users can create, view, and check off agenda items and action items, upload and step through presentations, type or draw annotations on top of slides, and mark entrance and exit of participants. Thus, not only are the artifacts and annotations themselves recorded, but all of the artifact interactions become indices into the meeting.

After a meeting is completed, the meeting records are automatically available for retrieval. The MeetingViewer, shown in Figure 1, integrates all of the meeting information based on time. The viewer uses a two-scale timeline for navigating a meeting, providing random access playback with finer-grained navigation on the lower timeline. The lower timeline shows a portion of the entire meeting, indicated by the black box focus region in the upper timeline. The timeline is painted with color-coded events as both a visual summary of the meeting and as an aid for navigation. The events currently captured by MeetingClient are people joining and leaving, agenda items being discussed, action items visited or created, and slides visited. However, we designed the timeline to include other events such as people speaking and keyword locations should the capture capabilities evolve. Users can control which of

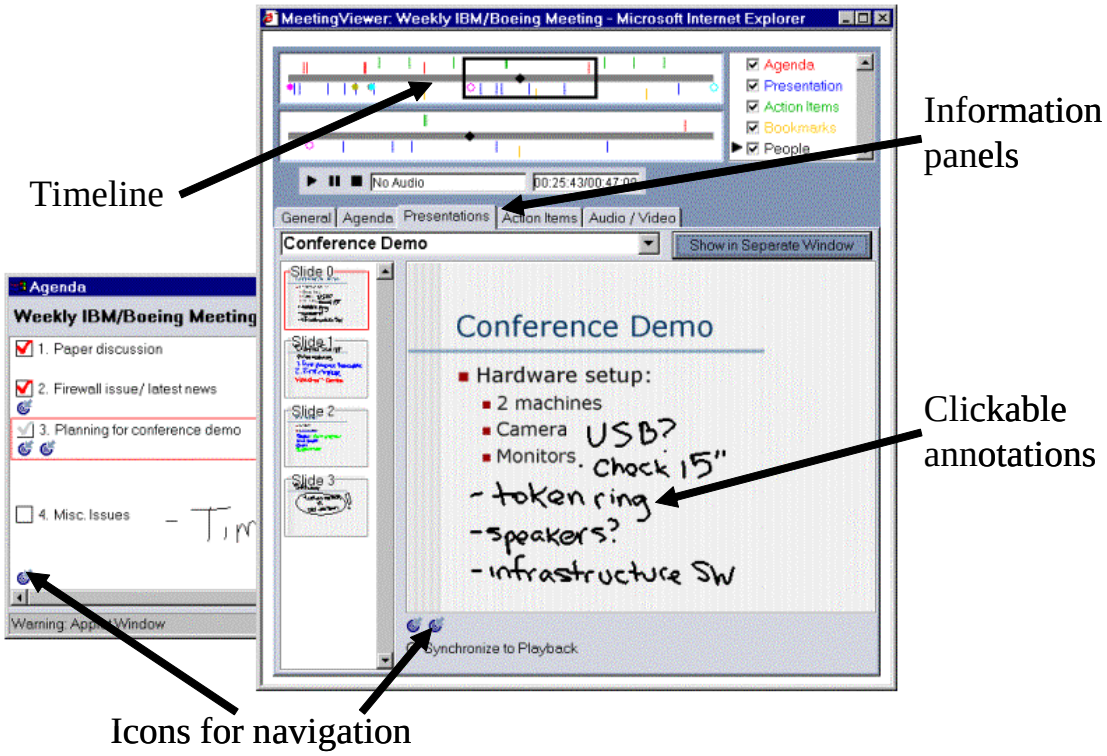

Figure 1. Screenshot of MeetingViewer.

these events they view and can use the events to find relevant portions within a meeting to playback. Additionally, users can click on the timeline or an event line, as well as drag the timeline scrub, to navigate the meeting. Playback of a meeting not only involves playing the audio and video, but also involves playback of all of the recorded events of a meeting such as slide visits and annotations.

The remainder of the meeting information is displayed on a series of tabbed panes for each of the objects related to the meeting, including descriptions and summaries of the meeting, agenda, presentations, action items, and video images. These panes are a very general approach for displaying a large amount of related information. However, to enable customized views, each pane can be opened in a separate window, moved and resized. This way, users can view any subset of the information they wish at once. Additionally, as TeamSpace evolves, we can easily add more meeting-related objects to this interface as another tabbed pane, such as documents that were reviewed or referenced during the meeting. We started with this rather general review interface so we can learn more about the types of information users need for various tasks and understand how to design task-oriented views that are simpler and more integrated in the future.

Users can navigate using the artifacts, with the agenda, action item and slide artifacts, using a small target-shaped icon beneath each object to move to the point in the meeting where that item was discussed. Users can also click on slide annotations, both ink and text, to jump to when those annotations were created. Annotations are grayed out if they have not been created at that point in the meeting, but are still clickable.

TeamSpace differs from previous meeting capture prototypes in its integration within a collaborative workspace, the capture of explicit interactions with artifacts, and the focus on a usable access system. While we did not push the envelope of technology, and we can imagine a great number of improvements and additions to Teamspace, we did succeed in creating a deployable, usable system.

STUDY DESCRIPTION

Our goal is to understand how people may browse, search, and utilize information in recorded meetings. One challenge in examining user access behaviors is that a real review may be a fairly rare occurrence, and thus difficult to observe. We used a controlled study, but as part of real adopted use, in order allow us to observe and question users searching for information in as realistic a situation as possible. For over two years, we regularly recorded the weekly meetings of our university research group, composed of about a dozen students and one faculty member. For a small number of those meetings, we asked group members, and occasionally a visitor to the group, to answer several questions regarding a previous meeting. Most of the subjects had attended the meeting, although several had not, or had missed a portion of the meeting. The

subjects were instructed to answer the questions as best they could, using MeetingViewer only if desired. Thus, the task was imposed, but the meetings and information contained in them were completely natural.

Subjects did occasionally remember the answers to questions, but most often used MeetingViewer to find the answers. Subjects were video-taped and asked to think aloud to help determine what memories and strategies they used to find information. All interaction with MeetingViewer was logged. After each review session, users were given a questionnaire asking them their impressions on the effectiveness of the searches and the usefulness of the different artifacts and aspects of TeamSpace.

The questions that subjects answered about the meetings varied greatly, dependent on the content of the meeting. While the questions were certainly not important to all subjects, we tried to represent realistic queries from the meeting. Most questions could not be fully answered without the audio and were likely not noted during the meeting by anyone, as few members of the group took personal notes. Questions ranged from those asking for simple facts, such as "When is Bob going to be out of town?" to those asking for summaries of entire discussions, such as "What is the overall recommendation for reshaping John's paper?" Questions were created by re-listening to the meeting and brainstorming a list of possible questions. Three or four questions were then chosen, attempting to balance the type of question, the topic of the question, and the possible ways to access the answers.

Data Collected

We evaluated 32 review sessions, involving 18 subjects and 10 meetings (some subjects reviewed more than one meeting.) Eighteen sessions were performed within one week of the recorded meeting, six within one month, and the rest more than one month later.

With the exception of one 4-hour retreat, all of the meetings lasted between 70 and 100 minutes, which was typical of the weekly group meetings. Audio was always recorded, but never video. All of the meetings had agendas, ranging from 2 to 6 items. Four meetings included one action item. Four meetings included prepared presentations, with two of those containing almost no additional notes or annotations. For the remainder of the meetings, one to three blank slides were used to take general notes. There was an average of 16.8 events captured for each meeting (not counting the participants joining and leaving) and an additional 21.2 text or ink annotations. These meetings statistics are typical of all of the meetings recorded over the years of deployment with this group.

Subjects remembered surprisingly few details from a meeting. Subjects only answered 15% of the questions from memory and attempted to use MeetingViewer to find the answer for the remaining questions. All subjects used MeetingViewer to help answer at least two of their questions in each session, and were not always able to find an answer. On 7% of the questions, subjects gave up their search and did not provide answers.

Task performance and behavior varied widely, both because of the different questions and meetings, and individual differences in what participants remembered and how they searched for answers. Users spent anywhere from 20 seconds to 15 minutes using MeetingViewer to answer a question, averaging 4:05 minutes with a median of 3:05 minutes. They played an average of 11:05 minutes of audio per session, using the timeline an average 43.5 times, and the meeting artifacts an average 5.6 times to navigate the recording. Subjects gave up on a question after anywhere from 1.5 to 12 minutes, averaging 5.5 minutes. Several gave up only after many minutes of searching without success, but others gave up quickly because they had no idea where to look and did not wish to waste time searching. These statistics show that despite relatively sparse indices, most subjects could and did find the information they were looking for. However, searching the meeting could also be time consuming and unsuccessful.

It is difficult to further characterize and understand the behavior of the participants with these simple statistics. Participants answered questions with different completeness or depth, some taking more time to find additional details. Subjects fluidly answered questions, finding an answer to one while looking for another, some taking the time to write in complete sentences, others with as few words as possible. The ways which subjects searched often differed. Some missed opportunities to use indices that would have made search easier. Several randomly searched and quickly got lucky, while others unfortunately jumped right over a desired audio segment.

ANALYSIS OF USER NAVIGATION

In order to gain a better understanding of the results and user behavior for each review session, we created a visualization of each subject's behavior over time, showing where in the meeting recording she was playing and when and how she navigated the playback. An example of this visualization is shown in Figure 2. The x-axis represents time, in minutes, elapsed in the review session. In the upper portion of each graph, the y-axis is the location of playback within the meeting record. Black dots denote when the audio was started, stopped, or moved. Lines indicate that audio was playing. The lower portion of the graph indicates what information was being viewed in MeetingViewer and what was used to navigate. In this case, the y-axis is divided into information categories. A line for a category indicates that the information panel was being viewed. For example, in Figure 2, the user was viewing both the agenda and the presentation for most of his review session. The timeline is always visible, so no line was drawn to indicate this. Dots indicate when an information panel was opened or closed, or when information on that panel was used to move the playback point in the meeting. In Figure 2, the presentation was used to navigate the meeting, as well as the timeline.

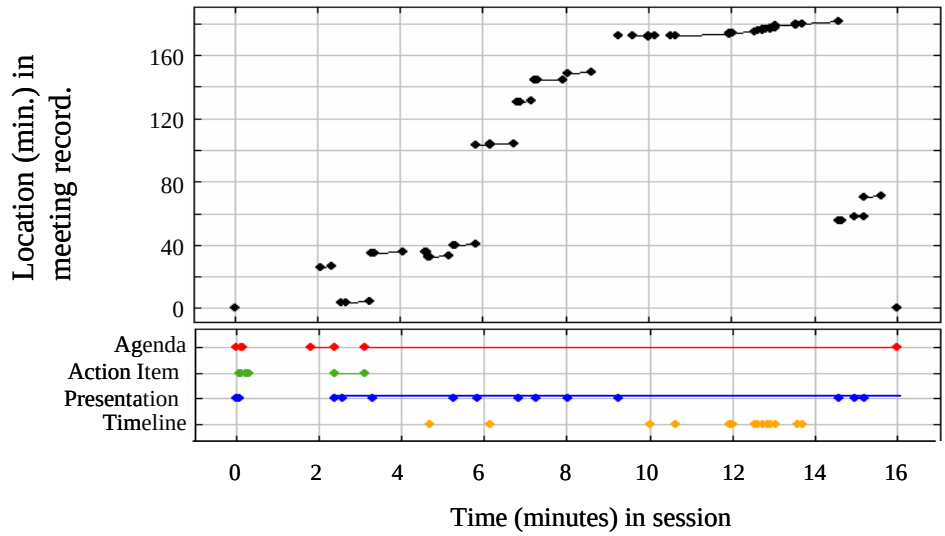

Figure 2. Review session for John.

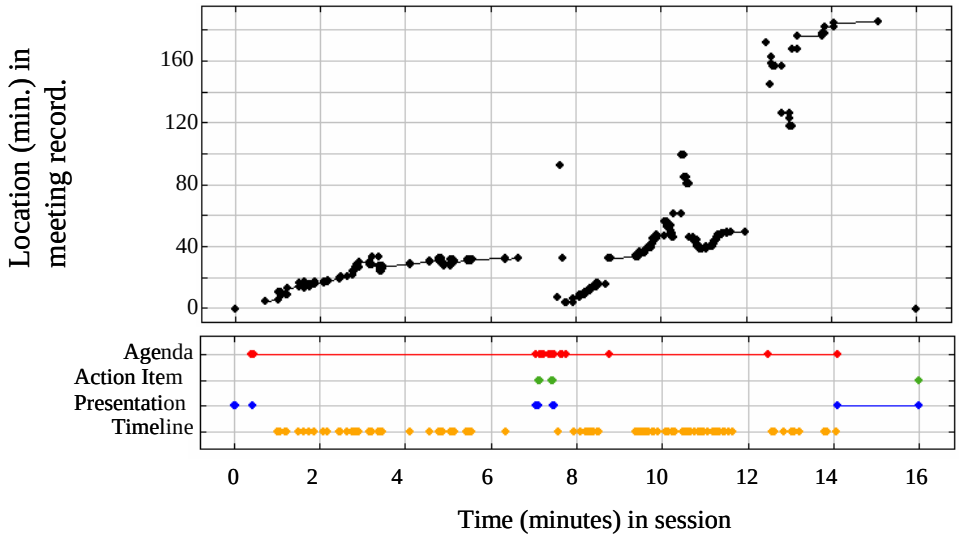

Figure 3. Review session for Bob.

With these visualizations, we were able to look for and identify patterns and differences in meeting navigation. Figure 2 shows "John's" review session for one meeting, while Figure 3 shows "Bob" reviewing the same meeting, answering the same set of questions. Both review sessions lasted roughly the same amount of time, yet their review behavior was very different. As the visualizations show, John frequently used the presentation, annotations in this case, to move through the meeting. On the other hand, Bob preferred the timeline and mainly navigated in this way. John's search was much more focused, and as such, his answers to the questions were more complete. These two visualizations highlight two distinct patterns – *jump* and *skim* – that show how using the artifacts as indices affected navigation.

Jump

Jump is using an artifact, such as an agenda item, slide annotation, or action item, as an index to move directly to a particular point in the meeting. These artifacts provided a semantic structure to the recording and related to the content of the discussion. In Figure 2, John uses the presentation, specifically annotations on a slide, to jump multiple times through the meeting. This is indicated by the dots on the presentation line, which correspond to jumps in meeting location in the upper portion of the graph. This navigation allowed him to largely ignore most of the meeting and only focus on the small segments he wished to hear. On average, users jumped 5.6 times during a review session, but with much variation. Users often started an information search with one or two jumps to find a segment of interest.

Using artifacts to jump to a particular point was easy for users to understand. When successful, jumping required very little interaction and was used whenever available. For example, even when the answer to a question was written as an annotation on the slide, subjects frequently clicked on that annotation to jump the playback point and listen to the audio while they wrote their answer, even though the audio was not required. This may indicate that audio would be consulted more frequently if doing so was only one click away while viewing other information.

However, applications need to make it clear how to perform the jump action. We had a usability issue with the navigation icon below each agenda, action item, and slide that confused users and resulted in fewer successful jumps than users desired. More importantly, the artifacts themselves did not show the flow of time through them. This was confusing when an agenda item or slide was visited multiple times throughout the meeting. While the timeline did show the order of events, matching this to the corresponding agenda items or slides was difficult, making the event lines on the timeline less useful. As an improvement, the icon for each agenda or action item could have instead been represented with a miniature timeline showing when during the meeting that item was relevant. Thus, linking the static artifact with its dynamic events, such as showing when it was the focus of attention, will improve navigation.

Skim

Skim is using the timeline to systematically navigate through the meeting record. This behavior was used when looking for a particular topic or segment of interest in the audio. Sometimes this behavior was also used to get a general idea of what was discussed throughout a portion of the meeting. This behavior is manifested by using the timeline repeatedly for small navigations in the meeting. For example, in Figure 3, Bob spends most of the review session skimming through portions of the meeting. Skimming was almost always done forward in time; users would only skim backwards if they realized that they had overshot an area of interest. Users often listened to the audio for a very short amount of time, usually about 1 to 3 seconds, before moving to the next point. The distance of skips varied both by person and by task. Some preferred smaller moves, about 10 seconds. Others skipped a minute at a time. A few participants used the event lines of the timeline to help them know where they might want to move, but most appeared to rely simply on time, and the audio, to help them know where they were.

Skimming could be very time consuming, especially if the area of interest was large. Users could spend more than 10 minutes skimming just to answer one question. However, this task was generally not frustrating as long as the user was making progress. While most people thought they had a good idea of the order of topics, they were often wrong, causing them to skip over a part of interest, or be looking in the wrong area completely. If this occurred, a user rarely wanted to re-skim that portion of the meeting and often gave up. Certain meetings seemed to be easier to skim than others, based upon the content and language of the discussion. The speaker's identity and the words used often gave a very quick indication of the topic. However, meetings with lots of informal comments, chitchat, long silences, and tangents seemed to be more difficult to skim. Users had to listen for longer to determine where they were in the meeting before skimming again. Users had different thresholds for time spent skimming. Some were more willing than others to spend time searching through a meeting before giving up, and even the same user could have different thresholds on different days. An additional confound was that if the user was making progress by skimming, she sometimes did not check for a useful index first, making the search longer and more difficult.

The timeline was extremely flexible; users could move any amount that they wished. Yet, this flexibility was often not necessary, and instead, introduced confusion. The user was required to judge the distance of his move based on the physical distance on the timeline, and the result of a click was not previewed ahead of time. This seemed to be more problematic the smaller the move. Several users requested a fast forward button that moved ahead a pre-set amount of time in order to facilitate skimming. Several other users requested ways to automatically skip useless audio, such as dead space or chitchat. This could sometimes decrease the amount of audio the user listens to before moving on, potentially improving skimming behavior.

Additional Patterns

While the jump and skim behaviors were the most frequent and critical we observed, and are greatly affected by the kind of indexing provided, we identified several other navigational patterns that also have implications for the design of access applications.

Scan

Scan involves quickly glancing at the information in the interface to get a feel for what is available. Users flipped through all of the information panels very quickly, spending only a few seconds on each. This behavior can be seen in the beginning of the session of Figure 2, where John scanned twice through the panels before he began any other review. In Figure 3, Bob also performed a scan at approximately 7 minutes into his review session, flipping through each panel multiple times. When a presentation was given, users similarly flipped through many or all of the slides using the slide thumbnails, again spending only a brief amount of time on each slide.

Usually at least once during the review session, averaging 1.4 times per session, users would scan to get a feel for the information in the interface. This scan usually included the empty video panel as well. Some users scanned at the

beginning of their session. However, scanning often occurred when the user was lost or frustrated.

Scanning was not a difficult task. The information panels were successful in that they were easy to understand and easy to extend. However, knowing that information even existed required going to that panel. On multiple occasions users would forget to check a particular panel, or forget what they had seen there. They thus missed using an index for quick access, and instead had to search the audio to find an answer. There will always be a tension between putting too much information on one page, and spreading out the information between panels. In this case, the sparseness of these meetings would have allowed us to group more information together more effectively. Task-oriented or customizable interfaces would help with these issues.

Honing

Honing could be considered a form of skimming, and the interaction is the same: using the timeline to skip through audio to find a particular segment. However, honing is more fine-grained, with a more focused goal. The general segment of interest has been discovered; the user basically knows where he is in the meeting. He just needs to find the exact point he is looking for. In this case, the user is skipping smaller amounts, usually 5 to 20 seconds of the audio, sometimes as little as a second or two.

We separate honing from skimming because the occurrence and implications for the two behaviors may be different. Honing occurs less often than skimming, and only when users are trying to find an exact piece of audio. A fast forward button would improve both behaviors, but the size of the desired move may be different. Yet, while honing may occur less often than skimming in this study, we observed that users were generally impatient and often moved ahead just a few seconds rather than listen to that audio to get to their desired point. This indicates that users may appreciate the ability to hone.

Replay

Replay is moving backward and playing a previously heard segment of the meeting record another time. Users frequently replayed the same small segment of audio multiple times to hear an exact statement. This sometimes occurred three or four times if the audio segment was difficult to hear or if the subject was exactly quoting a phrase as an answer. This behavior was generally performed using the timeline. However, if the audio segment started at a particular index, that index was used instead.

As with skimming and honing, the use of the timeline could introduce problems. The user had to determine how far back she wanted to move and the corresponding timeline location. Sometimes users did not skip back far enough, and sometimes they went too far and either had to hone again or listen to more audio than needed. Similar to improvements in skimming, users requested a reverse button to move back a pre-set amount to aid in replay.

Random

Random is a move, using the timeline, with no particular target. This is often indicated by several fairly large moves with no apparent pattern of navigation. Figure 3 shows Bob do several random moves around minute 10.5 and again at minute 13. Subjects stated that when they had no idea where to look, they sometimes randomly moved about on the timeline. One subject even said she used a binary search-like approach to narrow down the relevant area on the timeline. This behavior would usually not last long, as a subject would frequently find some segment of interest within a few clicks and begin skimming or honing. Randomly moving was also remarkably successful on a few occasions as a user happened across an answer without really knowing where she was looking.

Order of behaviors

While the order of these behaviors was greatly dependent on the user and the question, a prototypical, successful, search for information was as follows. The subject would start his use of MeetingViewer by looking at the agenda. If he saw an agenda item that corresponded to the topic he was looking for, he jumped to the beginning of that agenda item. If that agenda topic covered a long period of time, the subject would skim through the audio to find the more detailed topic he desired. Then the user would hone in on the exact answer and replay the answer once found. At some point, the subject scanned the information panels. He used an annotation on a slide to jump to a topic, and again skimmed and honed until the desired location was reached. In looking at our two previous examples, we see that John performed less skimming than usual while Bob performed more. A more balanced session can be seen in Figure 4, where “Jean” shows most of this prototypical sequence. She initially uses the agenda to find an area of interest, then uses the timeline to skim further. She finally visited the presentation towards the end of her session and used it to jump one final time.

USE OF INDICES

The most important measure of an access interface is how well it supports users in finding the information they want. And, as we have emphasized, the captured artifacts and indices are critical to supporting the browsing and searching for information. On the post-task questionnaire, subjects were asked, “How easy or difficult was it to find the information you were looking for?” The answer was “Moderately easy” in 57% of the questionnaires, “Not easy or difficult” in 23%, and “Moderately difficult” in 20%. Using a 5-point scale, this led to an average of 3.4, somewhere between “Moderately easy” and “Not easy or difficult.” Many of the difficult ratings were given when users spent a significant portion of their time skimming the interface.

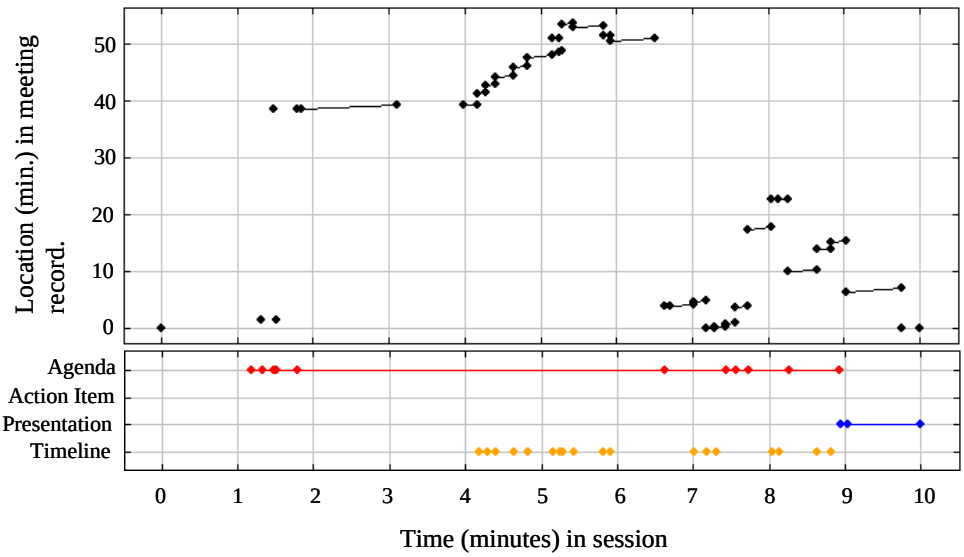

Figure 4. Review session for Jean.

Users clearly found information faster when they were able to use indices to find areas of interest, or even the exact desired location. The more indices we can provide, the more likely we are to improve users' searches for information. Every index we provided was used, including the rare action item. Over all sessions, the agenda was used to navigate 78 times, action items 3 times, slides 8 times, and annotations on slides 188 times. Indices that provide semantic structure of the recording were consulted first. The agenda was almost always viewed and used to start navigation. In the questionnaires, five subjects stated they wanted more agenda items, or more fine-grained topic changes, to provide additional structure. The more time between the activity and the review, the more likely users will need general, high-level structures to serve as an overview and reminder of what occurred in the meeting. This structure will, in turn, spark other memories of the content they are looking for and its context.

While high-level structure was useful for getting started, users desired other artifacts or indices that were directly connected to the content of the meeting. For example, the presentations and annotations contained useful content, and also structured the meeting record. On the questionnaire we asked which artifacts and events were useful, and the agenda and the presentation were almost always rated highly useful. In contrast, subjects rarely reported that the people joining and leaving events were useful. In feedback about what would facilitate their tasks, four participants wanted more notes to use, and 4 others requested just more indices in general. Overall, 11 subjects reported wanting additional indices in some form to help find information. In other words, subjects seemed to value all of the indices that related to the content of the meeting, and seemed to want as much of this as possible.

This study also shows that any type of index can and will help users find details within the recorded meeting. Anything that can be accurately captured is likely to be useful to someone at some point. Users did make use of the join and leave events, using the knowledge that what they were looking for occurred after a particular person entered or exited the meeting. One of the most requested improvements to MeetingViewer was speaker context. Four subjects suggested speaker identification and four others, speaker change information. Several others also wanted detection of silence and chitchat. Not only can this speaker context help users search based on who was speaking, they can also indicate the type of discussion occurring – such as a presentation with one speaker, versus a discussion with rapid switches between speakers. Speaker changes can also indicate topic changes during certain types of meetings, as different speakers address different topics. When the content-based indices have narrowed down the area as much as possible, these indices would then speed the process of skimming and honing on the desired details. In response, we are currently adding speaker change information to TeamSpace to investigate the use of such contextual indices more deeply. These and other forms of context-based indices are being actively researched in multimedia and information retrieval communities. [1, 12]

The other heavily requested addition was a searchable transcript or use of keyword spotting. This would allow keywords to function as a content-based index, narrowing down the areas of interest. A transcript could also eliminate the need for audio skimming, substituting text skimming instead, which would likely be faster. In this case, audio would only need to be played when the transcript was not sufficient to convey what was said, either because of recognition errors or because other verbal cues are not present. We did not use speech-to-text technology because of our low technology set-up with relatively poor audio quality. As this technology improves, we will be able to take advantage of these capabilities to improve access to

meeting content. However, as we demonstrated in other work on capturing knowledge acquisition meetings [10], we do not believe that a transcript will eliminate the need for content-based indices that structure the discussion. Instead, it will improve performance when content-based indices are not available.

In TeamSpace, we did not explore post-hoc or review-oriented indices, *i.e.* indices that were created after the meeting based on users' review of the recording. Several users also expressed a desire to bookmark areas of interest. However, this would only be useful if the user, or perhaps others, were going to later return to that portion of the meeting, which did not occur in this study. Users may also benefit from knowing where they have been in the recording. While not common, a few users did become lost and unintentionally revisit the same portion of the meeting multiple times. Indications of where the user has visited may lower these occurrences, and also help with replay. For example, if the user skips over a desired segment while skimming, he could focus on only the areas he had not visited yet.

One issue with any of these indices is that few of the users interacted with the capture application. This led to misunderstandings of what certain indices meant and how they were created. For example, users wanted the people events to represent speaker identification, which our system could not do, instead of the join and leave events captured by checking off participants in the capture tool. When a problem occurred, such as accidentally not selecting an agenda item during the meeting, users attributed the problem to the access interface instead of a capture issue. Improved labeling or a help interface could alleviate some of the confusion. However, conveying the meanings and potential problems with various indices will always be a challenge when the reviewer is not familiar with the capture capabilities and technology and how the events are generated.

SUMMARY OF DESIGN IMPLICATIONS

The results of our evaluation, including the navigational patterns and the use of various indices suggest improvements to TeamSpace and the design of access interfaces in general. The primary implication is that applications should try and provide as many content-based indices as possible. Session-related activities or contextual indices are generally non-persistent and simply mark when an activity was started or an event occurred. The study confirms our desire to provide a number of indices based on the artifacts in the meeting, which potentially describe what people are doing with objects and relate to the content of the discussion. Content-related indices help show the structure of the overall recording, were critical to quickly finding a portion of the meeting, and users wanted more of them. In other words, applications should support indices for jumping, and support skimming explicitly when necessary. While any and all indices are useful, capturing contextual indices such as speaker change that can help users infer content will also significantly aid navigation of the meeting record.

An application can further support replay with features to speed up skimming – by providing faster audio playback, by detecting and skipping over useless audio segments like silence or chitchat, and by providing fast forward or similar explicit navigational features within the interface. Additionally applications should have little or no delay when navigating the media stream to support quick skimming and honing. The interface should also indicate the flow of time through the artifacts, showing when each was relevant during the meeting. Finally, easy and flexible random access playback should always be provided to support unusual or unanticipated behaviors, like the random moves we frequently observed.

CONCLUSION

We have focused on the behavior of reviewing recorded meeting information in a controlled, yet realistic, evaluation. We have previously argued [5] that meeting capture applications should focus more around the artifacts of the meeting. That argument is confirmed by our observations of the importance of content-related indices to navigating the meeting record. Users rarely began a search thinking about contextual information from the meeting. Instead, they started with the topic of the information they were looking for, and then turned to context to aid their search from there. We would like to see more research on various methods for recording the content of the meeting through artifacts, and continued effort on recognizing such content through audio, video, and text analysis.

We also observed a number of specific navigational behaviors that have implications for the design of interfaces supporting review. Despite the decade and a half of meeting capture research, this is the most detailed evaluation of the behaviors of an access interface in this domain that we are aware of. Only the evaluations of eClass [2] and Tivoli [9] have provided comparable details, yet still did not discuss the navigation behavior at the same level of abstraction. Thus, our study has revealed even more details and specific interface issues that relate to such behaviors. However, the task supported in this evaluation was a search task, where users found a desired portion once and never returned. Thus, there are likely to be additional behaviors and issues that are important to other tasks. For example, Moran *et al*. observed a behavior they called salvaging in their evaluation of Tivoli [9], the high level behavior of finding, pulling out, and structuring interesting portions of the meeting record. However, many of the basic browsing and search behaviors we have identified are likely to be a part of any access of meeting information, and are thus important to understand and support.

In general, MeetingViewer was usable by the subjects during their task. Most were able to find answers to the questions, and many made very positive comments about

the interface. The interesting thing to note about MeetingViewer is that it cannot merely be usable. It must be "walk up and usable." Users were only ever given very brief training the first time they used the interface. And for some, this training was several years ago with only one or two subsequent uses per year.

While MeetingViewer was generally usable we found several issues that arose because a single user generally captured the meeting on a laptop instead of conducting the meeting using the capture application on a viewable whiteboard. Many users did not experience or view the capture functionality, and thus, did not always fully understand the review capabilities. These problems demonstrate the challenges in providing users an understanding of the capture capabilities, and the need to design features that support the desired information in the meeting.

Our experiences with TeamSpace also suggest a number of challenges that remain in creating useful and usable meeting capture applications. First, we believe that finding information within a meeting recording is still too time consuming for everyday use. We would likely not observe the same lengthy navigations in more natural use. Users still have issues in understanding how information is captured, what is available, and how to utilize it effectively. We feel that review interfaces are still relatively immature and not particularly innovative, our own included. Yet we still believe that automated capture technologies can be useful in many situations, and more research should focus on the review and use of recorded content. We are continuing to look at real uses and benefits of capture technology, in meetings and beyond.

ACKNOWLEDGMENTS

Special thanks to the original TeamSpace team of Werner Geyer, Shahrokh Daijavad, Ludwin Fuchs, and Steven Poltrock. Additionally, we thank the members of the Ubicomp research group at Georgia Tech for being guinea pigs of TeamSpace for several years, especially to those who were subjects in this study.

REFERENCES

1. Bett, Michael, Ralph Gross, Hua Yu, Xiaojin Zhu, Yue Pan, Jie Yang and Alex Waibel (2000). Multimodal Meeting Tracker. in the *Proceedings of RIAO 2000*, Paris, France.
2. Brotherton, Jason A. and Gregory D. Abowd (2004). Lessons Learned from eClass: Assessing Automated Capture and Access in the Classroom. *ACM Transactions on Computer-Human Interaction (TOCHI)*, **11**(2): 121-155.
3. Chiu, Patrick, Ashutosh Kapuskar, Sarah Reitmeier, and Lynn Wilcox (1999). NoteLook: Taking Notes in Meetings with Digital Video and Ink. In the *Proceedings of ACM Multimedia '99*, Orlando, FL.
4. Cutler, R., Y. Rui, A. Gupta, JJ Cadiz, I. Tashev, L. He, A. Colburn, Z. Zhang, Z. Liu and S. Silverberg, "Distributed meetings: a meeting capture and broadcasting system", in *Proceedings of the tenth ACM international conference on Multimedia*, Dec. 2002.
5. Geyer, W., H. Richter, and G. D. Abowd. Towards a Smarter Meeting Record – Capture and Access of Meetings Revisited, *Multimedia Tools and Applications*, Springer Science + Business Media, vol 27, no. 3, December 2005, pp 393-410.
6. He, Liwei, Jonathan Grudin and Anoop Gupta (2000). Designing Presentations for On-Demand Viewing. In the *Proceedings of Computer Supported Cooperative Work (CSCW'00)*, Philadelphia, PA, pp 127-134.
7. He, Liwei, Elizabeth Sanocki, Anoop Gupta and Jonathan Grudin (1999). Auto-Summarization of Audio-Video Presentations. In the *Proceedings of ACM Multimedia 1999*, Orlando, FL, pp 489-498.
8. Kazman, Rick, Reem Al-Halimi, William Hunt and Marilyn Mantei (1996). "Four Paragidms for Indexing Video Conferences." *IEEE Multimedia*, **3**(1): 63-73.
9. Moran, Thomas P., Leysia Palen, Steve Harrison, Patrick Chiu, Don Kimber, Scott Minneman, William van Melle, and Polle Zelweger (1997). "I'll Get That Off the Audio": A Case Study of Salvaging Multimedia Meeting Records. in the *Proceedings of Human Factors in Computing Systems (CHI'97)*, pp 202-209.
10. Richter, H., C. Miller, G. D. Abowd, and I. Hsi. "An Empirical Investigation of Capture and Access for Software Requirements Activities," in the *Proceedings of Graphics Interface 2005*, May 2005, pp 121-128.
11. Richter, H., G. D. Abowd, W. Geyer, L. Fuchs, S. Daijavad and S. Poltrock (2001). Integrating Meeting Capture within a Collaborative Team Environment. in the *Proceedings of UbiComp 2001*, Atlanta, GA.
12. Schultz, Tanja, Alex Waibel, Michael Bett, Florian Metze, Yue Pan, Klaus Ries, Thomas Schaaf, Hagen Soltau, Martin Westphal, Hua Yu and Klaus Zechner (2001). The ISL Meeting Room System. In the *Proceedings of Workshop on Hands-Free Speech Communication (HSC-2001)*, Kyoto Japan.
13. Stifelman, Lisa, Barry Arons, and Chris Schmandt (2001). The Audio Notebook: Paper and Pen Interaction with Structured Speech. In the *Proceedings of ACM Computer-Human Interaction (CHI 2001)*, Seattle, WA, pp 182-189.
14. Whittaker, Steve, Patrick Hyland, and Myrtle Wiley (1994). Filochat: Handwritten notes provide access to recorded conversations. In the *Proceedings of Human Factors in Computing Systems (CHI'94)*, pp 271-277.
15. Wolf, Catherine G., James R. Rhyne, Laura K. Briggs (1992). Communication and Information Retrieval with a Pen-based Meeting Support Tool. In the *Proceedings of Computer Supported Collaborative Work (CSCW'92)*.